

Big Data - Implementazione di una soluzione completa di analisi dei dati

Raccolta e analisi di Big Data

 A distanza



4 giorni (28 Ore)

Open : 2.490,00 € +IVA

WebCode: IT.27

Packaged in azienda : 7.250,00 € +IVA +10% di
Project Management (Quota riferita ad un gruppo
di 10 pax max)

Customized : Su richiesta

Comprendere le esigenze aziendali per fornire ai responsabili delle decisioni una serie di **indicatori affidabili e pertinenti** è il ruolo previsto degli specialisti dei Big Data.

Specificamente progettato per insegnare loro come implementare una soluzione Big Data in un ambiente **Hadoop** -una soluzione di punta per l'elaborazione di Big Data-, questo programma segue il percorso logico di **un progetto** di analisi dei dati. Implementazione di una soluzione completa di analisi dei dati che consentono l'organizzazione di un volume molto elevato di informazioni, alla produzione di script Pig e Hive che, convertiti in attività MapReduce, consentono di aggregare e filtrare i dati per poi analizzarli.

A chi è rivolto



Per chi

- Capi progetto
- Data Scientist, Data Analyst
- Sviluppatori
- Analisti e statistici
- Chiunque sia responsabile dell'implementazione operativa di un progetto Big Data in un ambiente Hadoop



Prerequisiti

La conoscenza di un linguaggio di programmazione e del linguaggio SQL agevola la partecipazione al corso

Programma

1 - Introduzione

- Obiettivi
- Schematizzazione del progetto
- Ecosistema e stack tecnologico
- Risultati attesi

2 - Ingestione massiccia di dati

- Descrizione
- Caratteristiche principali degli strumenti di assimilazione

- Soluzioni (package, cloud computing e Hadoop)
- Mettere a fuoco Apache NIFI
- Mettere a fuoco Apache KAFKA
- Ingestione di dati in streaming NIFI su KAFKA
- Realizzazione di un flusso di lavoro di “ingestion” di dati in streaming NIFI in HDFS

3 - Elaborazione in batch di Big Data

- Schema operativo
- Soluzioni software associate (MapReduce, Hive, Spark, ...)
- Script batch di Big Data
- Data Warehousing Big Data
- Creazione di tabelle partizionate, clusterizzate e/o transazionali
- Scrittura di script di aggiornamento per le diverse tabelle
- Analisi dei Big Data

4 - Elaborazione avanzata dei Big Data: machine learning

- L'ecosistema Spark
- Schema generale per la creazione di un modello ML
- Creazione di un modello di clustering dei dati
- Creazione di un modello di analisi predittiva supervisionato
- Arricchire un DW con Spark
- Salvataggio di un modello di machine learning
- Applicazione di un modello ML

5 - Archiviazione distribuita dei dati

- Principi delle basi di dati distribuite
- Soluzioni (NoSQL, NewSQL)
- Creazione di un database distribuito
- Inserimento di dati in un database distribuito
- Interrogazione di un database distribuito

6 - Automazione della catena di elaborazione batch

- L'orchestratore di Oozie
- Programmazione degli script HIVE
- Combinazione con script SPARK
- Creazione di un flusso di lavoro Oozie completo

7 - Elaborazione di dati massivi in flussi (streaming)

- Principio di funzionamento
- Soluzioni software
- Pre-elaborazione dei dati in streaming
- Realizzazione di flussi in un hub di streaming
- Uso di flussi in un hub
- Elaborazione avanzata dei dati in streaming (apprendimento automatico)

8 - Implementazione in un'architettura Big Data

- Approcci standard
- Realizzazione di una soluzione completa di elaborazione dati di tipo Lambda
- Realizzazione di una soluzione completa di elaborazione dati di tipo Kappa



Obiettivi del corso

- Avere le competenze tecniche necessarie per l'implementazione di analisi di Big Data
- Saper utilizzare gli strumenti di raccolta Open Source
- Essere in grado di scegliere la giusta soluzione di archiviazione dei dati per le specificità di un progetto (OLAP, NoSQL, GrafoDB)
- Esplorare la cassetta degli attrezzi tecnologici che costituisce Hadoop e il suo ecosistema e scoprire come utilizzare ogni "mattone" (MapReduce, HIVE, SPARK...)
- Imparare ad analizzare i risultati e a comprendere il significato dei dati estratti



Esercitazioni

- Una formazione molto **operativa** durante la quale si alternano le fasi di apporti teorici, scambi, condivisione di esperienze
- I partecipanti realizzano una catena di elaborazione Big Data (Batch e Streaming) su un **caso concreto**



Date 2026



Ultimi posti



Edizione garantita

dal 6 lug al 14 lug

- dal 6 lug al 7 lug
- dal 13 lug al 14 lug

dal 12 nov al 20 nov

- dal 12 nov al 13 nov
- dal 19 nov al 20 nov